Do We Have to Use the Power-law to Study War Size? And Does it Matter?

Miles D. Williams

Denison University[1]

2024-03-31

[1]Visiting Assistant Professor of Data for Political Research. Email: mdwilliams@denison.edu. Website: milesdwilliams15.github.io.

The Research Question

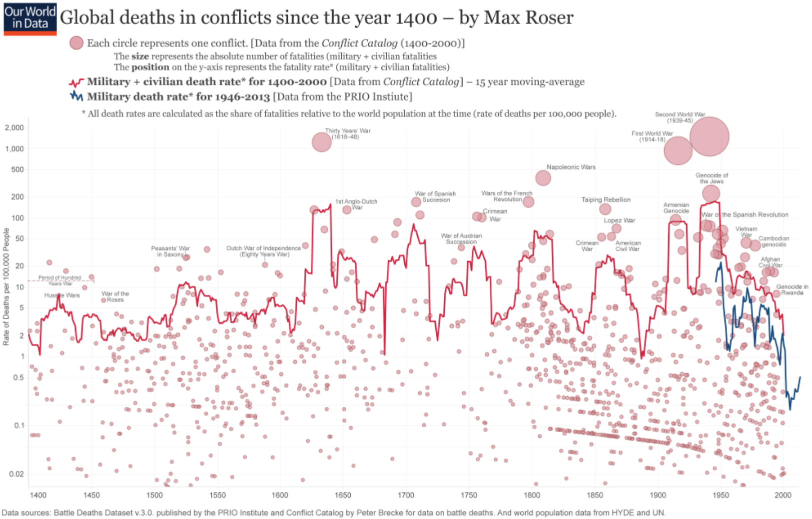
What is the best empirical model for studying trends in the sizes of international wars?

The Research Question

How should we go about evaluating competing empirical models of war size?

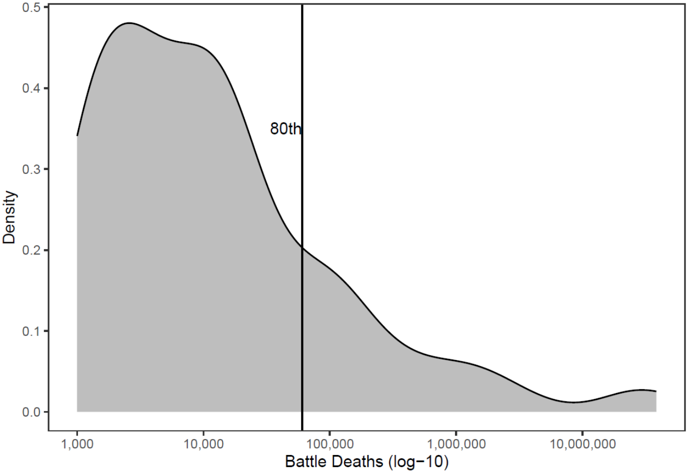
Why It Matters

Wars remain one of the few disasters (man-made or natural) that can snowball into hundreds thousands (if not millions) of deaths.



Why It Matters

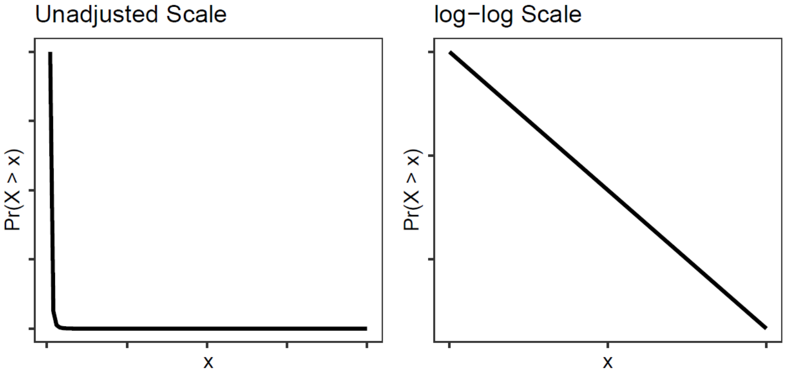
Their sizes have also been controversial to study statistically because of their heavily skewed distribution. The top 20% of wars account for 98.99% of battle deaths according to the CoW dataset.



Why It Matters

Because of its heavy-tail, the distribution of war deaths is conventionally modeled using the power-law, which holds that:

$$Pr(X > x) \propto x^{-\alpha}; \quad log[Pr(X > x)] \propto -\alpha log(x), \quad \forall \quad \text{large x}$$



Why It Matters

- Quantifying and explaining the sizes of wars is foundational to quantitative conflict research, but has since become niche.
- Interest has renewed with the emergence of a debate about the “long peace”
  - Some argue that trends in war size support the long peace (Cunen, Hjort, and Nygård 2020; Pinker 2011; Spagat, Johnson, and Weezel 2018; Spagat and Weezel 2020)
  - Others fail to identify statistical evidence for it (Braumoeller 2019; Clauset 2017, 2018)
- Differing data sources partly explain divergence on this issue, but the use of different statistical tools also accounts for it.

Why It Matters

If we want to move the debate about the long peace in fruitful directions, we need to take a step back and use best practices for model validation and comparison that are under-utilized in the peace science literature.

What I Did

- Used best practices outlined by Clauset, Shalizi, and Newman (2009) for fitting, validating, and comparing models fit to thick-tailed data
- Applied them to the popular CoW conflict series (Sarkees and Wayman 2010)
- Used three alternative specifications for war size

Methods

The procedure proposed by Clauset, Shalizi, and Newman (2009) is simple:

1. Fit models to data
2. Perform a simulation-based goodness-of-fit test using a distance statistic
3. Perform a non-nested likelihood ratio test for model comparison

Models

I tested out three alternative models:

1. The power-law (Braumoeller 2019; Cederman 2003; Cederman, Warren, and Sornette 2011; Cirillo and Taleb 2016; Clauset 2017, 2018; Spagat and Weezel 2020; Spagat, Johnson, and Weezel 2018)
2. The inverse Burr (Cunen, Hjort, and Nygård 2020)
3. The log-normal (Verbeeck et al. 2019)

Data

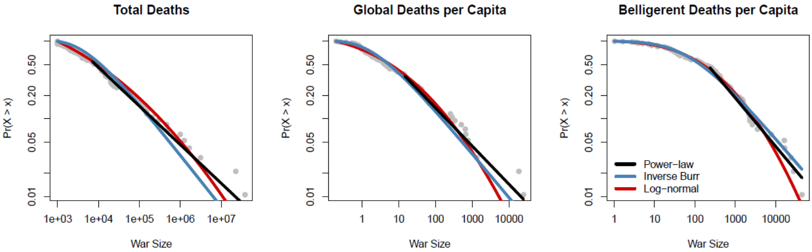
I performed this procedure with three measures of war size using the CoW international war dataset (N = 95 with coverage from 1816-2007):

1. Total Battle Deaths
2. Global Deaths per Capita (Total / Global Pop)
3. Belligerent Deaths per Capita (Total / Belligerent Pop)

Sources: Miller (2022); Sarkees and Wayman (2010); Singer, Bremer, and Stuckey (1972)

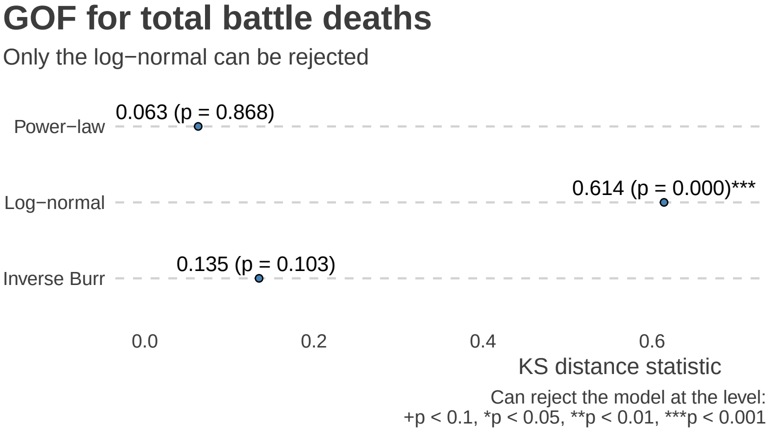
Model Fit

Step one is to fit the models to the data. Depending on how war size is measured, different models seem to fit better or worse.

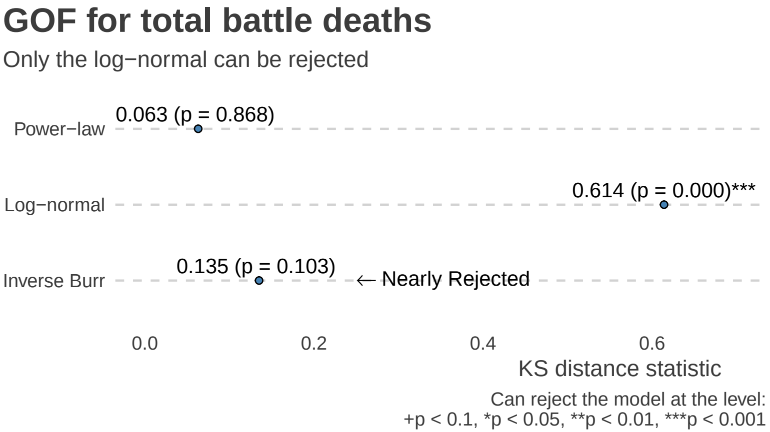


But we shouldn't just evaluate model fit by *look*. We need to engage in a replicable and transparent analysis to quantify goodness-of-fit.

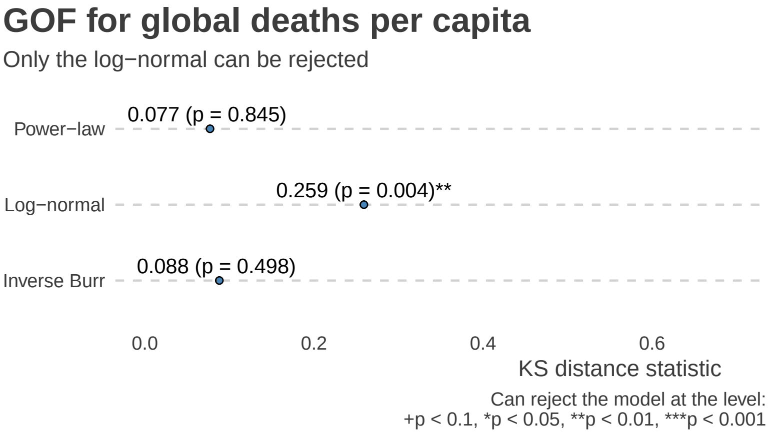
Test Goodness-of-Fit



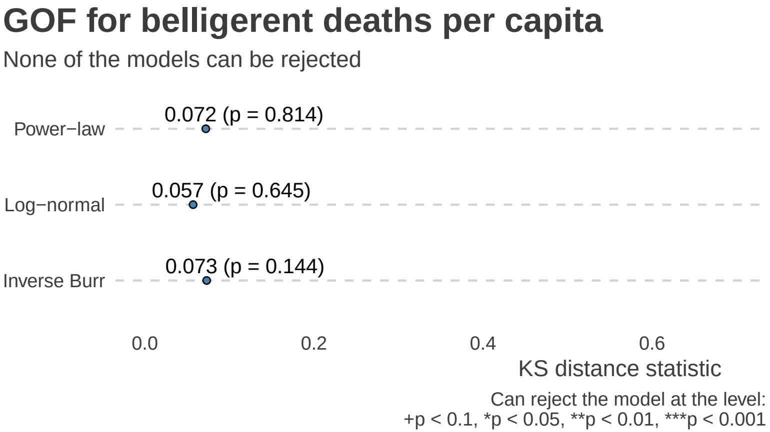
Test Goodness-of-Fit



Test Goodness-of-Fit



Test Goodness-of-Fit



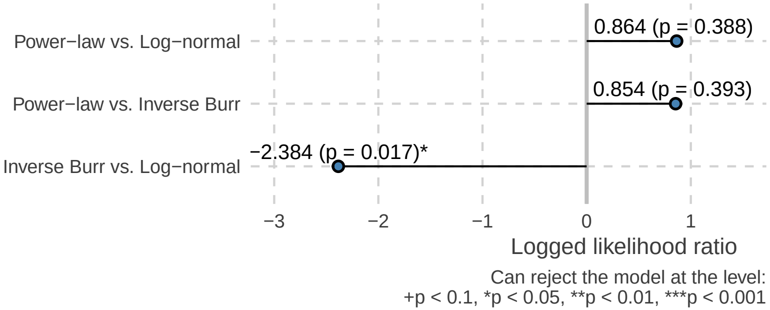
Model Comparison

Not all models can be rejected. When this is the case, we should then test whether any model performs statistically better than others.

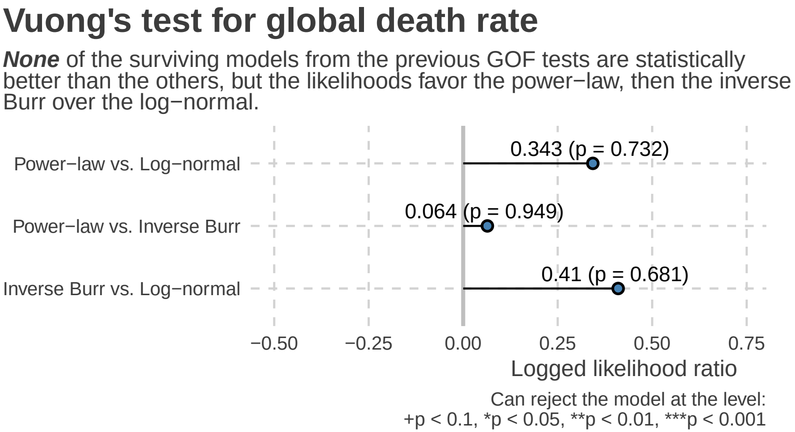
Model Comparison

Vuong's test for total deaths

None of the surviving models from the previous GOF tests are statistically better. But the log−normal, which was rejected earlier, is statistically ***better*** than the inverse Burr.



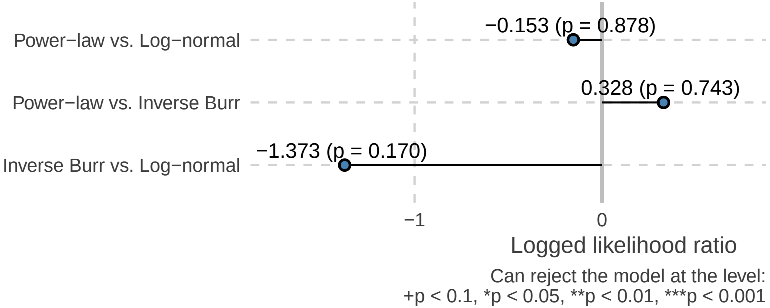
Model Comparison



Model Comparisons

Vuong's test for belligerent death rate

None of the surviving models from the previous GOF tests are statistically better than the others, but the likelihoods favor the log−normal, then the power-law over the inverse Burr.



Summary of Findings

All three models are justifiable:

- Total battle deaths: *power-law* and *inverse Burr* (barely)
- Global death rate: *power-law* and *inverse Burr*
- Belligerent death rate: *power-law*, *inverse Burr*, and *log-normal*

None of the surviving models is statistically better than the others.

Implications

How you measure war size determines which models may be a justifiable analysis tool.

Implications

- When multiple models are justifiable, choosing one model over another may influence results.
- Statistical precision and an ability to model covariates varies by models as well.

The measure and model may influence results

- I replicated prior studies on the long peace using the recently proposed 1950 cutpoint in conflict severity (Cunen, Hjort, and Nygård 2020; Spagat, Johnson, and Weezel 2018).
- Failed to find statistical support for the long peace, except when using the log-normal model and the belligerent death rate.
- When confronted with many justifiable choices, we should be wary of *p-hacking*.

Do we want to explain all the data?

- Best-practice with the power-law requires fitting a power-law slope to the most extreme observations in a distribution and ignoring smaller events.
- Data truncation can often exceed 50% of the data.
- This can hurt statistical precision and leave many data points unexplained.
- The inverse Burr and log-normal don't have this problem.

Do we want to justify regression analysis?

- The power-law (especially if its slope parameter is less than 2) makes regression analysis unjustifiable.
- In contrast, regression analysis with either the log-normal or inverse Burr is justifiable.

Conclusion

- This line of research is “in the weeds,” and the contribution is primarily methodological.
- But it is situated in a broader context of ongoing debates that have so far gone unresolved (and this is party because of the in-the-weeds factors to consider in this study).
- I don’t want to resolve any questions about the best model for war size or the long peace—but I do want us to work from a better methodological foundation in order to address these questions in more fruitful ways.

Thank you!

Miles D. Williams

Email: mdwilliams@denison.edu

Website: milesdwilliams15.github.io

References

Braumoeller, Bear F. 2019. *Only the Dead: The Persistence of War in the Modern Age*. New York: Oxford University Press.

Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *American Political Science Review* 97 (1): 135–59.

Cederman, Lars-Erik, T. Camber Warren, and Didier Sornette. 2011. "Testing Clausewitz: Nationalism, Mass Mobilization, and the Severity of War." *International Organization* 65 (4): 605–38.

Cirillo, Pasquale, and Nassim Nicholas Taleb. 2016. "On the Statistical Properties and Tail Risk of Violent Conflicts." *Physica A: Statistical Mechanics and Its Applications* 452: 29–45.

Clauset, Aaron. 2017. "The Enduring Threat of a Large Interstate War." Technical report. One Earth Foundation.

———. 2018. "Trends and Fluctuations in the Severity of Interstate Wars." *Science Advances* 4 (2): eaao3580.

Clauset, Aaron, Cosma Rohilla Shalizi, and M.E.J. Newman. 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.

Cunen, Céline, Nils Lid Hjort, and Håvard Mokleiv Nygård. 2020. "Statistical Sightings of Better Angels: Analysing the Distribution of Battle-Deaths in Interstate Conflict over Time." *Journal of Peace Research* 57 (2): 221–34.

Miller, Steven V. 2022. "peacesciencer: An r Package for Quantitative Peace Science Research." *Conflict Management and Peace Science*. http://svmiller.com/peacesciencer/.

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking Adult.

Sarkees, Meredith Reid, and Frank Wayman. 2010. *Resort to War: 1816 - 2007*. Washington DC: CQ Press.

Singer, J. David, Stuart A. Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War and Numbers*, edited by Bruce Russett. Beverly Hills, CA: Sage Publications, Inc.

Spagat, Michael, Neil F Johnson, and Stijn van Weezel. 2018. "Fundamental Patterns and Predictions of Event Size Distributions in Modern Wars and Terrorist Campaigns." *PloS One* 13 (10): e0204639.

Spagat, Michael, and Stijn van Weezel. 2020. "The Decline of War Since 1950: New Evidence." In *Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences*, edited by Nils Peter Gleditsch, 129–42. Springer.

Verbeeck, Cis, Emil Kraaikamp, Daniel F. Ryan, and Olena Podladchikova. 2019. "Solar Flare Distributions: Lognormal Instead of Power Law?" *The Astrophysical Journal* 884 (50): 1–16.